U.S. Army Research Institute
for the Behavioral and Social Sciences

Research Report 1795

Assessing Situation Awareness in Field Training Exercises

Michael D. Matthews
U.S. Military Academy

Scott A. Beal
U.S. Army Research Institute

September 2002

U.S. Army Research Institute for the Behavioral and Social Sciences

A Directorate of the U.S. Total Army Personnel Command

ZITA M. SIMUTIS
Acting Director

Technical review by

Jean L. Dyer, ARI
Chris Talcott, U.S. Military Academy

NOTICES

DISTRIBUTION: Primary distribution of this Research Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: TAPC-ARI-PO, 5001 Eisenhower Ave., Alexandria, VA 22333-5600.

FINAL DISPOSITION: This Research Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Research Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
September 2002	Final	July 2001 – March 2002

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Assessing Situation Awareness in Field Training Exercises	
	5b. PROGRAM ELEMENT NUMBER 62785
6. AUTHOR(S) Michael D. Matthews (U.S. Military Academy), Scott A. Beal (U.S. Army Research Institute)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER
	5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Behavioral Sciences and Leadership Engineering Psychology Program U.S. Military Academy West Point, NY 10996	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue. Alexandria, VA 22333-5600	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Research Report 1795

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*:

The Mission Awareness Rating Scale (MARS) and the Situation Awareness Behavioral Rating Scale (SABARS) are metrics designed to assess situation awareness (SA) among infantry soldiers and their leaders. MARS is a subjective self-assessment device and SABARS involves expert observer-controllers evaluating a target soldier on SA-related behaviors. The purpose of the current study was to field test both metrics in a field training exercise. Eight cadet platoon leaders and eight cadet squad leaders participating in summer cadet field training at the U.S. Military Academy completed the MARS instrument and received SABARS evaluations from observer-controllers following the completion of an assault mission. Results indicated that platoon leaders rated their SA higher than did squad leaders on the MARS instrument and that higher-order SA was rated as more difficult than lower- order SA . SABARS ratings did not differ as a function of leader position, but the global SABARS SA item was a strong predictor of ratings of individual performance. Finally, SABARS was rated by the observer-controllers as easy to use and relevant to assessing SA in the field. Both MARS and SABARS show promise of applicability to assessing SA in field settings.

15. SUBJECT TERMS Situation awareness (SA), decision making, SA measures, team SA, individual SA, shared SA, small unit leadership, workload

SECURITY CLASSIFICATION OF			19. LIMITATION OF	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	OF ABSTRACT Unlimited		Dr. Michael D. Matthews 845-938-3696

Research Report 1795

Assessing Situation Awareness in Field Training Exercises

Michael D. Matthews
U.S. Military Academy

Scott A. Beal
U.S. Army Research Institute for
the Behavioral and Social Sciences

Infantry Forces Research Unit
Scott E. Graham, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

September 2002

Army Project Number
2O262785A790

Personnel Performance
and Training Technology

FOREWORD

Contemporary military operations are characterized by a high operational tempo and a large volume of time-sensitive information that leaders and soldiers must quickly and accurately assess in order to successfully complete mission objectives. Situation awareness (SA) is a concept closely linked to effective decision making. A vital step in developing programs to enhance SA, or to evaluate the impact of new procedures or technology on SA, is the development of SA measures that are both psychometrically sound and user-acceptable.

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) began a series of research initiatives in SA by hosting the Infantry Situation Awareness Workshop in September of 1998. Since then, several approaches to measuring SA among infantry soldiers and their leaders have been developed and reported by ARI researchers. The two measures tested in the current study were previously tested in a virtual environment. The data reported here extend the application of these SA measures to a field training exercise conducted at the U.S. Military Academy as a part of cadet summer training. Because the bulk of infantry training occurs in field settings, establishing the applicability of SA measures in the field is a key step in developing SA metrics that can be used by researchers and trainers. To be useful, field SA measures must be acceptable to potential users and relatively unobtrusive, in addition to being valid and reliable.

This research represents the continued evolution of SA metrics developed by ARI. The results were briefed to the Department of Military Instruction, U.S. Military Academy, West Point, New York, on 1 May 2002.

MICHAEL G. RUMSEY
Acting Technical Director

ASSESSING SITUATION AWARENESS IN FIELD TRAINING EXERCISES

EXECUTIVE SUMMARY

Research Requirement:

This research is part of an ongoing effort by the U.S. Army Research Institute, Infantry Forces Research Unit, Fort Benning, Georgia, to develop effective measures of situation awareness (SA). In particular, there is a need for SA measures that are easy and practical to administer yet provide useful SA information to researchers and trainers.

Procedure:

The Mission Awareness Rating Scale (MARS) and the Situation Awareness Behavioral Rating Scale (SABARS) were tested on eight cadet platoon leaders and eight cadet squad leaders engaged in a platoon-level field training exercise during Operation Highland Warrior at the U.S. Military Academy. MARS is a self-assessment instrument, and SABARS is a rating completed by expert observer-controllers. Each platoon attacked an enemy command and control center then secured a nearby village.

Findings:

Platoon leaders rated their SA higher than did squad leaders on MARS. Higher-order SA was rated more difficult than lower-order SA, and workload measures were rated as more demanding than SA content. SABARS ratings did not differentiate as a function of leader position, but the global SABARS SA rating was a strong predictor of individual performance and decision making. Moreover, observer-controllers rated SABARS as easy to use, useful in providing feedback to the leaders they observed, and as a potentially viable training assessment tool.

Utilization of Findings:

The results extend previous validations of MARS and SABARS from the virtual environment to a field training setting. Although limited by a relatively small number of participants, the results suggest both MARS and SABARS have potential for assessing infantry SA in the field training environment.

ASSESSING SITUATION AWARENESS IN FIELD TRAINING EXERCISES

CONTENTS

Page

INTRODUCTION 1

METHOD 5

Overview 5
Participants 5
Instruments 5
Procedure 7

RESULTS 8

DISCUSSION 13

REFERENCES 16

APPENDIX A A-1

APPENDIX A: MARS INSTRUMENT A-1

B: SITUATION AWARENESS RATING SCALE B-1

List of Tables

Table 1. Mean and Standard Deviations of Responses to MARS Items by Platoon Leaders and Squad Leaders 9

2. MARS ANOVA Results 9

3. Correlation Matrix of MARS Items Leader Performance Indicators 12

4. Summary of Regression Analyses of MARS Items Predicting Leader Performance Indicators 12

CONTENTS (continued)

Table 5. Summary of Regression Analyses of Global SABARS.....13
SA Rating Predicting Leader Performance Indicators

List of Figures

Figure 1. Mean MARS ratings on Content subscale items as a function 10
of leadership position.

2. Mean MARS ratings on Workload subscale items as a function 10
of leadership position.

3. Mean MARS ratings for the Content and Workload subscales as.....11
a function of leadership position.

Assessing Situation Awareness in Field Training Exercises

Infantry leaders and their soldiers face a daunting task in performing their missions effectively. Soldiers perform missions in difficult terrain and weather. They may be deployed with little advance notice to a region of the world with which they are totally unfamiliar. Fatigue, noise, and stress combine to make the perception of key aspects of their environment difficult. In addition, soldiers engage an intelligent enemy who will actively attempt to deceive them in an attempt to accomplish their own mission. The infantry mission is by definition a team-based operation. The leader and soldiers must know where their fellow soldiers are located, their strength and combat readiness, and their physical and emotional readiness to perform the mission. Added to this, the Army is implementing sophisticated information technology systems down to the level of the individual soldier, thus adding further to the soldier's cognitive workload. To be effective, soldiers must be able to rapidly assess a situation, understand its meaning, and make valid predictions about what is likely to occur in the future. This ability defines the concept of situation awareness (SA).

Situation awareness is a construct closely linked to decision making. The most relevant model of SA for Army applications was described by Endsley, Holder, Leibrecht, Garland, Wampler, and Matthews (2000). This model, based on Endsley's (1988) conception of SA, specifies three levels of SA. Level 1 SA involves the perception of features critical to mission success. Level 2 SA is the understanding or comprehension of the meaning of the mission-critical features. The ability to make projections about what events are likely to occur in the immediate future defines level 3 SA. Establishing good SA is viewed, according to this model, as a critical precursor to effective decision making.

The development of effective and operationally relevant SA measures is vital in SA research. Until recently, little systematic effort had been reported in developing SA metrics in the infantry domain. Endsley et al. (2000) reviewed SA measurement approaches, and evaluated them for applicability to the infantry environment. According to Endsley et al., SA measurement approaches can be divided into four types: (a) process indices; (b) direct measures; (c) behavioral measures; and (d) performance measures. Process measures include eye movements, communications, and verbalizations. Direct measures include objective measures such as on-line probes, "freeze" probes, and subjective measures based on self and observer ratings. Behavioral measures involve inferring SA from specific behaviors on specific subtasks, such as "time to make a response (verbal or non-verbal) to some event, and correct or incorrect SA as identified from soldier verbalizations and appropriateness of a given behavior for a particular situation" (Endsley et al., 2000, p. 80). Performance measures are based on tactical performance during missions or exercises.

Each type of SA measure has strengths and weaknesses and may provide different sorts of information. For example, direct subjective SA measures have an advantage of assessing an individual's personal level of SA, are easily administered, and are relatively unobtrusive to collect. However, soldiers may not know what information they are unaware of and their judgments may be influenced by self-assessments of their own or their unit's performance. Similarly, direct objective SA measures, such as the situation awareness global assessment technique (SAGAT, see Endsley, 1995), have the advantages of providing more objective, and less biased estimates of SA, but are relatively obtrusive and require considerable prior analyses to develop valid measurement protocols. Given that different SA measurement approaches may provide different types of information or be more acceptable in certain situations, a measurement strategy utilizing multiple SA approaches is desirable.

Based on Endsley et al.'s (2000) analysis, Strater, Endsley, Pleban, and Matthews (2001) developed three SA measures specifically designed to measure SA among infantry small unit leaders and tested them during platoon level missions in a virtual environment. In this study, experienced and inexperienced platoon leaders led three squad leaders and computer generated forces through four missions in a virtual environment focusing on military operations in urban terrain (MOUT). The first measure was a SAGAT protocol modified to reflect SA requirements for infantry platoon leaders in MOUT missions. The second, the Situation Awareness Behavioral Rating Scale (SABARS), was a direct subjective measure utilizing expert observers to rate the platoon leaders on behaviors linked to SA in this context. The last measure developed was the Participant Situation Awareness Questionnaire (PSAQ), a subjective SA measure.

The SAGAT procedure involved periodically freezing action during the virtual MOUT scenarios, and then administering probe questions to the platoon leaders. These questions were based on a SA requirements analysis completed previously (Strater et al., 2001). A list of 21 queries were generated that indexed information critical to SA for small unit leaders in missions of this type. These queries corresponded to the three levels of SA defined by Endsley (1988). In the virtual environment, unlike the real world, the "ground truth" (i.e., the actual status of these factors, which can be verified in virtual simulations and carefully conducted field experiments) is known. Participants whose responses match the ground truth criteria, therefore, presumably possess higher SA than those whose responses fail to correspond to ground truth, thus providing an objective index of the individual's level of SA. Results showed that the SAGAT procedure differentiated SA as a function of the experience level of the platoon leaders. It also was sensitive to the MOUT scenario. Experienced platoon leaders showed better SA for enemy information such as location and strength, while less experienced platoon leaders focused their attention more on the status of their own forces.

The SABARS also showed promise for measuring SA. Based on the SA requirements analysis described above (Strater et al., 2001), 28 behaviors and actions linked to SA in MOUT missions were identified. Observer/controllers (O/Cs) closely watched each platoon leader work through the four scenarios. The O/Cs then rated the platoon leaders on a five-point scale with respect to how well they performed these behaviors. Like SAGAT, the SABARS data differentiated SA as a function of the experience level of the platoon leaders. Specifically, experienced platoon leaders were more likely to gather relevant information, follow procedures, and to focus on the big picture more than less experienced platoon leaders.

The PSAQ, which consisted of three questions dealing with workload, performance, and awareness of the situation as it evolved, showed the least promise for measuring SA. No significant effects were found for experience level, scenario, or the interaction between these two variables (Strater et al., 2001, p. 26). The PSAQ did not have questions specific to SA components and therefore may have been too general to be sensitive to subtle differences in SA.

The results reported by Strater et al. (2001) suggest that both SAGAT and SABARS hold promise for measuring SA and for use as tools in estimating the impact of new technology or operational procedures on small unit leader SA. These measures may also provide a basis for evaluating strategies designed to enhance SA. Further testing of both of these measurement approaches in other mission types and in field training exercises are necessary to fully evaluate their potential utility in SA research, development, and training.

Because SAGAT is a direct objective measure of SA, it might seem to be the preferred method for measuring SA. In simulations and other tightly controlled settings it may indeed represent the most desirable approach. However, much of the training in the Army is done in the field, and Army leaders and trainers are often reluctant to interrupt the flow of an exercise to allow measurement of what they may view as elusive psychological constructs. In these settings, the SABARS or a similar approach might be more desirable to the extent that it is less obtrusive. However, SABARS is labor intensive in the sense that O/Cs or other highly experienced personnel must be dedicated to observing a particular leader over the course of an exercise in order to provide meaningful ratings of SA related behaviors.

Thus, in some instances a subjective SA assessment procedure may prove useful. In Strater et al. (2001) the subjective measure, the PSAQ, did not provide useful assessments of SA. However, numerous other direct, subjective SA measures exist and have been used in a variety of domains (see Endsley et al., 2000). Recently, McGuinness and Foy (2000) reported the development of a subjective SA measure that can be easily tailored to a variety of settings or domains. McGuinness and Foy refer to this instrument as the Crew Awareness Rating Scale (CARS). This instrument has been applied to a variety of tasks and

contexts. The CARS consists of two sets of four questions. The first three questions of each set correspond to the three levels of SA defined by Endsley (1988), i.e., perception, comprehension, and projection. The fourth question of each set deals with how well the respondent identifies goals for the situation he or she is in. Moreover, the first set of four questions pertain to assessing SA content, for example, how well the respondent thinks he or she understands the situation. The second set of questions addresses workload, for example, how much mental effort is required to achieve understanding in a given situation. The assessment of workload is an important aspect of SA. There could be situations in which a person has high levels of SA, but most of their attentional capacity is required to achieve that level of SA. This would leave little mental workspace left to allocate to other, perhaps equally critical processes. This sort of subjective measure could prove useful in evaluating the impact of new technologies on SA and mission performance.

Matthews, Beal, and Pleban (2002) describe the initial testing and validation of a subjective SA measure based on CARS. The newly developed measure was designed to be consistent with the nature of the infantry mission. Because of this focus, this instrument was referred to as the Mission Awareness Rating Scale, or MARS. This research was based on an experiment designed to assess four different approaches to simulating night vision goggles in a virtual environment. Sixteen enlisted soldiers, working in four-man teams, completed four MOUT missions in a virtual environment. After each mission, participants completed a MARS questionnaire. The results indicated that MARS significantly and robustly differentiated among the four means of simulating night vision goggles, and that the MARS data were consistent with other objective and subjective indexes designed to assess the four different approaches to night simulations.

The two SA measurement approaches most suitable for field application, SABARS and MARS, have both been used in the virtual environment. The purpose of the current study was to test both of these measures in a field exercise. Because the majority of infantry training occurs in the field, and not in simulations, it is important to evaluate these instruments during a field exercise. This research was conducted during summer field training at the U.S. Military Academy. Both MARS and SABARS were used to assess the SA of cadet platoon and squad leaders engaged in an infantry exercise. Because platoon leaders had a broader view of the overall mission it was expected that platoon leaders would have higher SA than squad leaders as assessed by both MARS and SABARS. Both instruments should be predictive of outcomes conceptually related to SA, such as decision making and performance. With respect to MARS, self-assessment of SA should show that higher-order SA (e.g., projection) is more difficult than lower order SA (e.g., identification).

Method

Overview

The missions reported here were part of a larger exercise known as Operation Highland Warrior. Operation Highland Warrior is part of cadet summer training in which cadets execute a series of infantry missions over the course of several days. Organized as a battalion level exercise, Operation Highland Warrior centers around the theme of a deployment to the country of Cortina, where the battalion engages in peace enforcement operations. Rising seniors are selected to command the battalion, its companies, and platoons. Rising juniors are appointed to the role of squad leaders. The role of private is played by rising sophomores. The purpose of the exercise is to develop leadership skills in cadets. The current research focused on platoon level exercises.

Participants

The participants were 16 cadets. Eight were platoon leaders, and eight were squad leaders. The platoon leaders were rising seniors. They ranged in age from 21 to 24. One was female, and none had prior enlisted experience. The squad leaders were rising juniors. They ranged in age from 20 to 23, all were male, and none had prior enlisted service.

In addition, six infantry officers and four infantry noncommissioned officers, serving as O/Cs during the exercise, completed SABARS evaluations of the participants. Six of the O/Cs rated two platoon or squad leaders, and the remainder rated one. The officers were two majors and four captains, and had between eight and 22 years of active duty experience. The noncommissioned officers included three sergeants (E-5's) and one staff sergeant (E-6). Their active duty experience ranged from four to 13 years.

Instruments

Mission awareness rating scale (MARS). MARS (Appendix A) is based on the Crew Awareness Rating Scale (CARS) described in depth by McGuinness and Foy (2000). The instrument consists of two subscales. One assesses SA content (Content Subscale) and the other assesses SA workload (Workload Subscale). Each subscale consists of four questions that address the three levels of SA as defined by Endsley (1988) – identification, comprehension, and prediction. In addition, a fourth question deals with how well mission goals can be identified. For the Content Subscale, the four questions require the respondent to rate how well they can identify, comprehend, predict, and decide in the given mission. The Content questions are referred to in this report as Content-Identify, Content-Comprehend, Content-Predict, and Content Decide. The four Workload Subscale questions require the respondent to indicate how much mental effort is required to identify, comprehend, predict, and decide in the given mission. The Workload questions are referred to in this report as

Workload-Identify, Workload-Comprehend, Workload-Predict, and Workload-Decide. All questions were rated on a four-point scale.

Situation awareness behavioral rating scale (SABARS). The SABARS measure (Appendix B) consists of 28 questions relevant to infantry missions. The SABARS items elicited ratings from the O/Cs on how well the platoon or squad leader exhibited behaviors consistent with acquiring and disseminating SA information during the exercise. Since SA actually refers to an individual's internal representations of elements in the environment (perception, comprehension, and projections), it is important to note that the SABARS measure does not rate actual SA, but rather outward actions that indicate a greater likelihood of good internal representations. SABARS ratings were on a five-point scale ranging from "very poor" (1) to "very good" (5). A "not applicable/can't say" response was also available.

Leader performance rating. After completing the SABARS, each O/C was asked to rate his platoon or squad leader on four performance measures. Each question was rated on a five-point Likert-scale with responses ranging from "far above average" to "far below average." The four questions were:

1. "The performance of the platoon or squad as a whole on this mission was"
2. "The platoon or squad leader's decision making during this mission was"
3. "The platoon or squad leader's ability to work effectively with members of the unit during this mission was"
4. "I would rate the overall performance of this platoon or squad leader as"

SABARS evaluation. In addition, each O/C was asked four questions about the use of SABARS in assessing SA and leader performance. These questions were rated on a five-point Likert-style scale, ranging from "strongly disagree" to "strongly agree." The four questions were:

1. "SABARS included questions important in assessing situation awareness for small infantry teams"
2. "SABARS was easy to use"
3. "My ratings on SABARS could be used to give useful feedback to the leader on his or her mission performance"
4. "Providing a way for O/Cs to give trainees feedback on situation awareness is an important goal for improving training"

Procedure

Each platoon began its mission by being airlifted by helicopter to a staging area approximately one kilometer from an "enemy" command and control installation and a small village. The command and control installation consisted of a heavily fortified two-story structure with a large satellite dish adjacent to it. The command and control site was occupied by a heavily armed and firmly entrenched enemy force of seven soldiers. The exercise called for the American platoon to first attack and secure this installation, then enter and secure the nearby MOUT village.

The MOUT site was located about 75 meters east of the command and control installation. This site consisted of three concrete block buildings, including one that was two floors high. The MOUT site was occupied by several role players portraying the part of civilians on the battlefield. The first author played the role of the Sheriff of Cortina, and confronted the platoon leaders with a variety of requests and demands when they entered the village. Three to five other role players representing citizens of the village were present. Their job was to interact with entering soldiers, beg for food, protection, or medical assistance and, in doing so, to require the platoon's leaders to make decisions concerning the disposition of the civilians and their requests. In addition, two "enemy" soldiers, dressed as civilians, mingled with others in the village. They had hidden firearms and booby traps, and part of the platoon's mission was to identify these individuals and capture them. The sheriff and other legitimate occupants of the village had identification cards to show the American forces if asked to do so.

After arriving at the landing zone, the platoon took between four and six hours to devise a plan to approach and attack the command and control installation, then to enter and secure the village. The enemy force occupying the command and control installation consisted of highly experienced, active duty infantry soldiers from Fort Drum. The attack on this highly fortified and well-defended position was extremely difficult and tiring to the platoon, and they typically suffered many casualties. The attack, once initiated, lasted an average of 20 minutes. After completing this phase of the mission, the O/Cs conducted a brief after action review, then regrouped and the platoon approached the MOUT site.

When the platoon entered the MOUT site, the sheriff approached the platoon leader, asked several questions, and made demands for food and medical supplies for its inhabitants. Meanwhile, the platoon leader and the squad leaders engaged in actions to secure the village. This involved identifying enemy soldiers disguised as civilians and locating hidden arms or booby traps. This phase of the operation required an average of 30 minutes to complete.

Upon completing the MOUT phase of the mission, another after action review was conducted by the O/Cs. The experimenter approached the head O/C

and asked him to identify which O/C had observed the platoon leader. This O/C was asked to complete a SABARS evaluation on the platoon leader. In addition, the experimenter asked the head O/C which squad leader had been most involved in the mission, and which O/C had most closely observed that squad leader. The experimenter then asked the identified O/C to complete a SABARS evaluation on the identified squad leader.

As soon as the second after action review was completed, the experimenter asked the platoon leader and the squad leader to complete the MARS instrument. The forms required less than five minutes to complete.

A total of eight platoons executed the missions described here. Their structure was similar to active duty dismounted infantry platoons, consisting of three squads, a platoon sergeant, and a radio telephone operator (RTO) in addition to the platoon leader. Only one platoon per day was assigned this mission. Thus, MARS and SABARS data were collected on eight platoon leaders and eight squad leaders. Upon completing this phase of the mission, the platoons were transported by truck to their company area, where they bivouacked and prepared for subsequent missions.

Results

MARS

Table 1 shows the means and standard deviations for the eight MARS questions for the platoon leaders and squad leaders. Also, ratings for each of the Content and Workload questions were summed to form a Content Subscale and a Workload Subscale. Table 1 also shows the means and standard deviations of these. A 2 (position; platoon leader x squad leader) by 2 (subscale; Content x Workload) by 4 (question; identify x comprehend x predict x decide) ANOVA was performed and is summarized in Table 2.

Overall, there was a significant effect of position, with platoon leaders showing higher self-ratings of SA than squad leaders. The value of eta squared for this difference was .26, indicating a moderately strong effect size. There was also a significant effect for subscale (eta squared = .28) and item (eta squared = .17), with the Workload Subscale showing lower SA than the Content Subscale. There were no significant two-way interactions, but the three-way interaction between position, subscale, and item was significant (eta squared = .18).

Figure 1 shows the ratings for the four Content items as a function of position, platoon versus squad leader. The squad leaders rated all four items as more difficult than did the platoon leaders. The squad leaders rated identifying critical mission-related cues as the most difficult task. The platoon leaders rated identifying as somewhat more difficult than either comprehending or predicting, but rated deciding as the most difficult. Figure 2 shows the ratings that platoon and squad leaders made for the four Workload items. The squad and platoon

leaders rated identifying similarly. However, the squad leaders rated the other three questions – understanding, predicting, and deciding – as progressively more

Table 1.
Mean and Standard Deviations of Responses to MARS Items by Platoon Leaders and Squad Leaders

MARS Item	Position	N	Mean	Standard Deviation
Content-Identify	Platoon Leader	8	1.75	.25
	Squad Leader	8	2.62	.18
Content-Comprehend	Platoon Leader	8	1.62	.18
	Squad Leader	8	2.00	.19
Content-Predict	Platoon Leader	8	1.62	.18
	Squad Leader	8	2.12	.23
Content-Decide	Platoon Leader	8	1.88	.30
	Squad Leader	8	2.25	.25
Content Subscale	Platoon Leader	8	1.72	.19
	Squad Leader	8	2.25	.11
Workload-Identify	Platoon Leader	8	2.25	.31
	Squad Leader	8	2.12	.23
Workload-Comprehend	Platoon Leader	8	1.50	.19
	Squad Leader	8	2.38	.26
Workload-Predict	Platoon Leader	8	2.12	.30
	Squad Leader	8	2.50	.19
Workload-Decide	Platoon Leader	8	2.12	.23
	Squad Leader	8	2.75	.25
Workload Subscale	Platoon Leader	8	2.00	.21
	Squad Leader	8	2.44	.15

Table 2.
MARS Data ANOVA Results

Source	Mean Square	df	*F*	*p*	Eta Squared
Position	7.51	1,14	4.99	.04	.26
Subscale	1.76	1,14	5.41	.04	.28
Subscale x Position	.07	1,14	.26	.22	.01
Item	.86	3,42	2.84	.05	.17
Item x Position	. 09	3,42	.30	.82	.02
Subscale x Question	.34	3,42	1.25	.31	.08
Subscale x Item x Position	.86	3,42	3.14	.04	.18

difficult. In contrast, the platoon leaders ratings indicated similar perceived difficulty for each item with the exception of comprehend, which they rated as relatively easy.

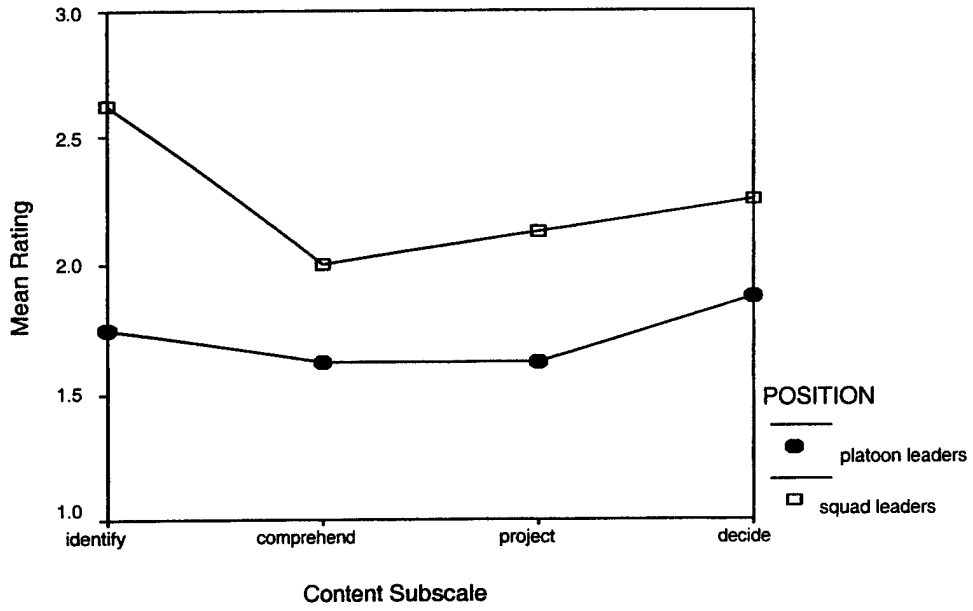

Figure 1. Mean MARS ratings on Content subscale items as a function of leadership position.

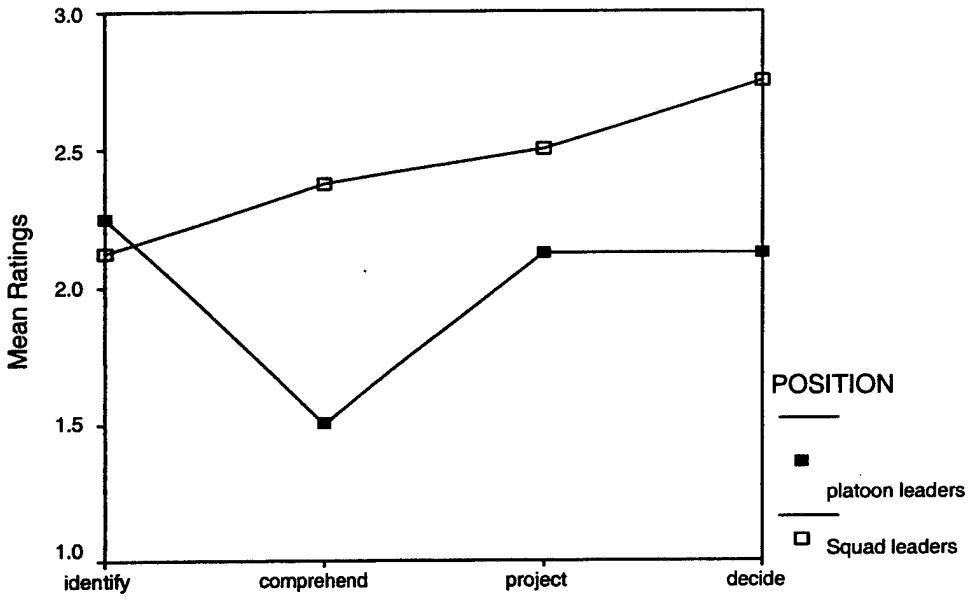

Figure 2. Mean MARS ratings on Workload subscale items as a function of leadership position.

Figure 3 shows the mean Content and Workload subscale ratings as a function of position. The main effects of position and subscale are reflected with squad leaders rating SA as more difficult than did the platoon leaders, and both groups rating SA Content as less difficult than SA Workload.

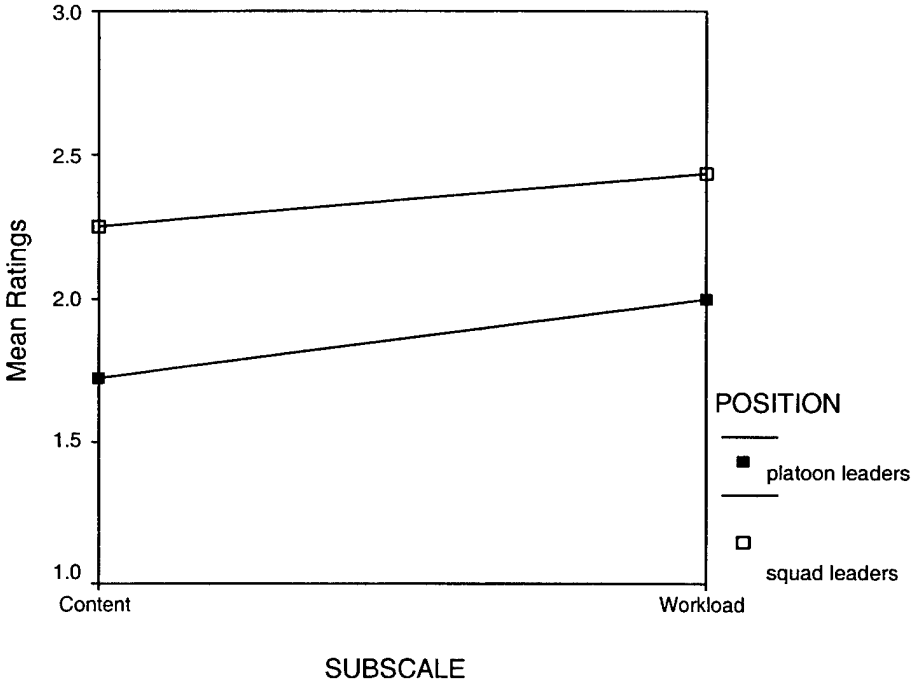

Figure 3. Mean MARS ratings for the Content and Workload subscales as a function of leadership position.

The correlation matrix of the eight MARS items is shown in Table3. Because of the small N, a relatively large correlation coefficient is needed to reach statistical significance. Of 28 possible correlations, eight were statistically significant. These were Content-Identify and Content-Comprehend (.58), Content-Comprehend and Content-Decide (r = .51), Content-Decide and Workload-Identify (r = .67), Content-Comprehend and Workload-Decide (r = .73), Content-Predict and Workload-Decide (r = .57), Content-Decide and Workload-Decide (r = .66), Workload-Identify and Workload-Predict (r = .51), and Workload-Comprehend and Workload-Decide (r = .53). Four other correlations showed *p* levels between .05 and .10. These were Content-Decide and Workload-Comprehend (r = .46), Content-Identify and Workload-Decide (r = .45), Workload-Identify and Workload-Decide (r = .45), and Workload-Predict and Workload-Decide (r = .50).

Data from the 2 MARS subscales were entered into a stepwise linear regression and used to predict the four general performance ratings of each platoon and squad leader obtained from the O/Cs. Table 4 summarizes the regression results. The multiple R's for the four equations ranged from .23 for the leader's ability to work effectively with others to .48 for the leader's decision

making ability. Although these values are relatively high, none reached statistical significance, presumably because of the relatively small N used in the study.

Table 3.
Correlation Matrix of MARS Items

	Content-Comp	Content-Predict	Content-Decide	Workld-Identify	Workld-Comp	Workld-Predict	Workld-Decide
Content-Identify	***.58***	.34	.32	.17	.38	.39	.45
Content-Comp		.32	***.51***	.26	.45	.16	***.73***
Content-Predict			.30	.20	.26	.25	***.57***
Content-Decide				***.67***	.46	.33	***.66***
Workld-Identify					.14	***.51***	.45
Workld-Comp						.41	***.53***
Workld-Predict							.50

Statistically significant coefficients (P. < .05) are in ***bold italics.***

Table 4
Summary of Regression Analyses of MARS Items Predicting Leader Performance Indicators

Criterion*	R	R Square	df	F	Sig.
1	.44	.20	2,13	1.59	.24
2	.48	.23	2,13	1.92	.19
3	.23	.05	2,13	.35	.71
4	.46	.21	2,13	1.70	.22

*Criterion Questions:
(1) "The performance of the platoon/squad as a whole on this mission was"
(2). "The platoon/squad leader's decision making during this mission was"
(3). "The platoon/squad leader's ability to work effectively with members of the unit during this mission was"
(4). "I would rate the overall performance of this platoon/squad leader as"

SABARS

SABARS ratings for platoon leaders and squad leaders were compared with an independent groups *t* test. Employing the Bonferroni correction, the two groups did not differ on any of the 28 comparisons. Because of a large number of missing observations and a small N, factor analyses and multiple regression with a large set of predictors was inappropriate. However, a global SA rating

(question 28) from SABARS was obtained for 15 of the 16 participants. This rating was used to predict the four criteria the O/Cs rated each leader on. Table 5 shows the outcome of these regressions. The regressions indicated a strong predictive relationship between the global SA rating and three of the four criteria, with only unit performance failing to reach statistical significance

Table 5
Summary of Regression Analyses of Global SABARS SA Rating in Predicting Leader Performance Indicators

Criterion*	R	R Square	df	F	Sig.
1	.49	.24	1,13	4.12	.06
2	.57	.27	1,13	6.25	.03
3	.85	.72	1,13	34.05	.00
4	.71	.50	1,13	12.93	.00

*Criterion Questions:
(1) "The performance of the platoon/squad as a whole on this mission was"
(2). "The platoon/squad leader's decision making during this mission was"
(3). "The platoon/squad leader's ability to work effectively with members of the unit during this mission was"
(4). "I would rate the overall performance of this platoon/squad leader as"

An important component in assessing the value of a rating scale is its acceptability to users. The O/Cs rated the SABARS on four questions. The first question was "SABARS included questions important in assessing situation awareness for small infantry teams." On a five-point scale, the mean response was 4.06, or "agree." The O/Cs also agreed that "SABRS was easy to use," with a mean rating of 3.94. They also agreed, with a mean response of 4.12, that "My ratings on SABARS could be used to give useful feedback to the leader on his or her mission performance." Finally, the O/Cs provided a mean rating of 4.25 to the question "Providing a way for O/Cs to give trainees feedback on situation awareness is an important goal for improving training."

Discussion

The results show that MARS differentiated SA between squad and platoon leaders conducting a training mission in a field environment, supporting the original hypothesis concerning MARS. This is consistent with the view that platoon leaders should have a broader picture of the mission, and therefore better SA. In partial support of the hypothesis, squad leaders (but not platoon leaders) rated higher-order SA (prediction and decision making) as more difficult than lower-order SA items (identification and comprehension). This is consistent with SA theory (Endsley et al., 2000). The SA Workload items were rated as more demanding than SA content items. Matthews et al. (2002) reported that MARS differentiated between four ways of simulating night vision goggles in a virtual environment. Thus, it appears that MARS is sensitive to different independent variables (i.e., ways of simulating night vision goggles and platoon

vs. squad leader), and is applicable to different settings (i.e., virtual vs. field). The ability of MARS to detect differences across experimental manipulations and settings suggests it may have general utility as an effective and user-acceptable measure of SA. The validity of MARS is supported because MARS data, in both the current and past studies, are consistent with theory-derived hypotheses.

The interpretation of SABARS data was hampered by the small N and large amount number of missing observations. None of the 28 items differed between squad and platoon leaders. However, the global SA rating from the SABARS was a strong predictor of leader performance, decision making, and working effectively with others. Importantly, SABARS was rated favorably by the O/Cs with respect to ease of use, importance in assessing SA, as a potential source of feedback to a small unit leader, and in overall importance of providing small unit leaders feedback on their SA. It took an average of five minutes to complete, and the O/Cs posed very few questions while completing the ratings. Usability is an important criterion in evaluating the practical utility of a field assessment instrument. SABARS appears to be viewed by users in a positive light.

SABARS seems to be a promising way of assessing SA in the field. In future tests the specific SABARS items should be carefully checked to verify that they correspond to the behaviors and actions required in the specific scenario or mission being tested. It would also be useful to provide raters with training and familiarization with the instrument prior to utilizing it in the field. The SABARS questions used in the current study were originally developed for platoon leaders operating in a virtual environment (Strater, et al., 2001). While many of the behaviors or actions required by platoon leaders are also executed by squad leaders, some probably are not. Future tests should focus on a specific mission and echelon.

Two factors limit the generalizability of the results. First, the platoon and squad leaders in the current study were U.S. Military Academy cadets. The overarching purpose of Operation Highland Warrior was to provide a demanding field-based leadership experience for the cadets involved. Thus, this was the first infantry based field exercise that they had ever participated in. The second limiting factor was the small number (eight platoon leaders and eight squad leaders) included in the sample. To evaluate more fully the psychometric properties and practical utility of these two SA measurement approaches, it is necessary to assess a larger number of experienced soldiers and their leaders in a more realistic field training exercise.

Although MARS and SABARS have now been tested in both virtual and field settings, additional research is needed to establish their reliability and validity. For example, although MARS appears to have differentiated the "SA" of platoon leaders and squad leaders in the current experiment, and also among different ways of simulating night vision goggles in a virtual environment

(Matthews et al., 2002), it is not yet clearly established that the instrument in fact measures SA. It is necessary to collect objective SA data simultaneously, perhaps in the form of a SAGAT measure, along with MARS in order to verify that MARS is, in fact, tapping into SA and not some other characteristic.

In conclusion, the current report shows that MARS and SABARS can be employed in a field training setting. Measures of these types are more acceptable to field trainers than the better validated but more obtrusive SAGAT protocols of the type more appropriately used in simulations (e.g., Strater et al., 2001). Additional research is needed to validate MARS and SABARS, but it appears that these measures show the potential of offering psychometrically sound and user-acceptable measures of SA in field settings.

References

Endsley, M.R. (1988). Design and evaluation for situation awareness enhancement. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 97-101). Santa Monica, CA.

Endsley, M. R., Holder, L. D., Leibrecht, B. C., Garland, D. J., Wampler, R. L., & Matthews, M. D. (2000). Modeling and measuring situation awareness in the infantry operational environment. (Research Report 1753). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Matthews, M. D., Beal, S. A, & Pleban, R. J. (2002). Situation Awareness in a Virtual Environment: Description of a Subjective Measure. (Research Report 1786). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McGuinness, B., & Foy, L. (2000). A subjective measure of SA: The Crew Awareness Rating Scale. In D. B. Kaber and M. R. Endsley (Eds). Human performance, situation awareness and automation: User centered design for the new millennium. Atlanta, GA: SA Technologies.

Strater, L. D., Endsley, M. R., Pleban, R. J., & Matthews, M. D. (2001). Measures of platoon leader situation awareness in virtual decision making exercises (Research Report 1770). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Appendix A
MARS Questionnaire

Name _____ Date _____

Position (check one) ____ Platoon Leader _____ Squad Leader

Class: _____ 2002 _____ 2003

Mission Awareness Rating Scale (MARS)

Instructions. Please answer the following questions about the mission you just completed. Your answers to these questions are important in helping us evaluate the effectiveness of this training exercise. Check the response that best applies to your experience.

The first four questions deal with your ability *to detect and understand important cues present during the mission.*

1. Please rate your ability to ***identify*** mission-critical cues in this mission.

___ very easy – able to identify all cues
___ fairly easy – could identify most cues
___ somewhat difficult – many cues hard to identify
___ very difficult – had substantial problems identifying most cues

2. How well did you ***understand*** what was going on during the mission?

___ very well – fully understood the situation as it unfolded
___ fairly well - understood most aspects of the situation
___ somewhat poorly – had difficulty understanding much of the situation
___ very poorly – the situation did not make sense to me

3. How well could you ***predict*** what was about to occur next in the mission?

___ very well – could predict with accuracy what was about to occur
___ fairly well – could make accurate predictions most of the time
___ somewhat poor – misunderstood the situation much of the time
___ very poor – unable to predict what was about to occur

4. How aware were you of ***how to best achieve*** your goals during this mission?

___ very aware – knew how to achieve goals at all times
___ fairly aware – knew most of the time how to achieve mission goals

___ somewhat unaware – was not aware of how to achieve some goals
___ very unaware – generally unaware of how to achieve goals

The last four questions ask how *difficult* it was for you to detect and understand important cues present during the mission.

5. How difficult – in terms of mental effort required - was it for you to ***identify*** or detect mission-critical cues in the mission?

___ very easy – could identify relevant cues with little effort
___ fairly easy – could identify relevant cues, but some effort required
___ somewhat difficult - some effort was required to identify most cues
___ very difficult – substantial effort required to identify relevant cues

6. How difficult – in terms of mental effort – was it to ***understand*** what was going on during the mission?

___ very easy – understood what was going on with little effort
___ fairly easy – understood events with only moderate effort
___ somewhat difficult – hard to comprehend some aspects of situation
___ very difficult – hard to understand most or all aspects of situation

7. How difficult – in terms of mental effort – was it to ***predict*** what was about to happen during the mission?

___ very easy – little or no effort needed
___ fairly easy – moderate effort required
___ somewhat difficult – many projections required substantial effort
___ very difficult – substantial effort required on most or all projections

8. How difficult – in terms of mental effort – was it to decide on ***how to best achieve*** mission goals during this mission?

___ very easy – little or no effort needed
___ fairly easy – moderate effort required
___ somewhat difficult – substantial effort needed on some decisions
___ very difficult – most or all decisions required substantial effort